REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
August 1998

3. REPORT TYPE AND DATES COVERED
Final Report

4. TITLE AND SUBTITLE
Visualization Methods for the Exploration of High Dimensional Data

5. FUNDING NUMBERS
DAAH04-94-G-0267

6. AUTHOR(S)
Edward J. Wegman

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)
George Mason University
Fairfax, VA 22030

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
ARO 32850.17-MA

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

19981229 119

13. ABSTRACT *(Maximum 200 words)*

This project brought together elements of virtual reality technology (immersive technology) and elements of scientific visualization to address issues of high dimensional spatial and volumetric visualization problems. Applications of such methods were to the visualization of any spatially or volumetrically distributed data with multivariate data characterization at each pixel or voxel. A military application was, for example, characterization of mine fields. This research accomplished under this funding has a very strong dual usage character.

continued on reverse side

14. SUBJECT TERMS

15. NUMBER IF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT

18. SECURITY CLASSIFICATION OF THIS PAGE

19. SECURITY CLASSIFICATION OF ABSTRACT

20. LIMITATION OF ABSTRACT

Visualization Methods for the Exploration of High Dimensional Data: Final Report

Edward J. Wegman

Army Research Office
DAAH04-94-G-0267

Duration: 7/15/94 to 9/30/97

George Mason University
Center for Computational Statistics
157 Science-Technology Building #2
Fairfax, VA 22030

Abstract

This project brought together elements of virtual reality technology (immersive technology) and elements of scientific visualization to address issues of high dimensional spatial and volumetric visualization problems. Applications of such methods were to the visualization of any spatially or volumetrically distributed data with multivariate data characterization at each pixel or voxel. A military application was, for example, characterization of mine fields. This research accomplished under this funding has a very strong dual usage character.

We developed several visualization methodologies based on the concept of a mathematical grand tour and based on recursive, data-adaptive pattern recognition. These visualization methods are embedded in immersive display systems which include a head-mounted display system and a stereographic projection system. Research included tasks oriented toward developing approximation models for general two- and three-dimensional geometric structures, visualizing spatial and multivariate multivariate data, integrating high performance parallel and graphics computers, and developing psychometric comparisons for stereo visualization.

Computer and virtual reality resources were largely in place so that there was little or no latency in the development of this project. An added bonus is the leverage obtained with the impending installation of an Intel Paragon supercomputer at George Mason University. This project was very effective producing 26 technical reports including two Ph.D. dissertations and 30 published papers.

Introduction

This proposal focused on the alliance of elements of virtual reality technology and elements of scientific visualization to address issues of mine detection and related spatial and volumetric visualization problems. By virtual reality or virtual environments, we mean an immersive visual and audio technology such that experimenter has little or no awareness of the real environment. For our purposes of data visualization, this is intended as a focusing device so that the experimenter has a heightened sense of awareness of the problem at hand and, thus, can concentrate in a natural way his or her full mental resources. We believe that this singular focus is particularly valuable in adversarial settings in which human lives are at stake. Moreover, we believe that the immersive technology, because of its three dimensional aspects, will be less fatiguing than other types of human-computer interfaces, current technology limitations notwithstanding.

Much of currently fashionable work on scientific visualization has been focused on rendering of flow fields arising from combustion or meteorological applications, molecular, atomic or subatomic particle dynamics, and other settings modeled with partial differential equation models. Our focus had been, in contrast, on data representation, exploratory data analysis and model building using high performance computer graphics. It was our intention to continue the development of our graphics and visualization tools in the context of immersive technology.

Virtual Reality, Immersive Technology and Scientific Visualization

The development of virtual reality as a graphics construct began in the mid-seventies with attempts to develop more realistic flight simulation. The standard paradigm was to have training pilots look at a visual display screen. The effect, as might be expected, was that these pilots were aware that they were looking at a visual display screen and, consequently, had a comparatively strong sense of unreality about the whole exercise. An approach to overcoming this problem has become known as virtual reality. The intent of the virtual reality construct is to create a more intimate sense of involvement (originally in the flight simulation), i.e. a sense of immersion. Virtual reality constructs can be implemented in several ways. The goal is to recreate via computer the items that the subject's senses might sense in reality. Visually, this would mean replacing a flat display screen with binocular stereo views which would change along with the motion of the subject's head. That is, if the subject looked down, he or she would see the view which is below. If the subject looked backwards, he or she would see the view in back. This is accomplished with a headgear containing miniature binocular video displays. There is an audio component which can be achieved in the same headgear by incorporating binaural headphones. Finally, one would want to create a tactile component which can be achieved by incorporating instrumented gloves. The technology is now becoming commercially available, but has been demonstrated for several applications including the flight simulation application as well as for an interesting application giving a sense of liberation to handicapped individuals.

The major new thrust in statistical data analysis and visualization is to combine scientific visualization, exploratory data analysis and virtual reality into a new technology for exploring data. There is a history of several years of experimentation with 3-D visualization and many aspects of this tool are in hand. Data analysis has seen movement of two-dimensional paper-based graphics in the seventies to the computer based three-dimensional color graphics of the eighties. Three-dimensional graphics have been achieved by kinematics displays where motion gives the depth cues and by stereoscopic displays. We have used both red-green stereo and polarized-light stereo to good effect. However, as satisfying as these displays are, they clearly represent the paradigm of a scientific investigator looking at a screen. What we are doing now is to leave this scenario and immerse the scientific data analyst in a data world. Allowing the analyst to literally move about in the data world, to fly around the data, to look at different portrayals of the data, to make a turn and literally turn into another dimension seamlessly. We would imagine a direct manipulation setting where an analyst can actually grab hold of a data point in this virtual world and see the effect of moving it around. We would imagine a progressive disclosure setting in which, as the analyst approached a data point in this virtual data world, the numerical value or other attributes of the data point would come into focus. What we are proposing is to create a data analysis environment where the scientist could quite literally explore the data much the same way that an oceanographer with scuba gear or a submersible could explore the sea.

Indeed, this undersea exploration analogy is a good one. To understand the potential impact, one can imagine an oceanographer who wishes to explore the ocean. On the one hand, the oceanographer could create a remotely piloted vehicle with a television

and light system and watch the exploration process on a video monitor within a darkened room in his ship. This is to a large extent what computer-based visualization does now. A second model might be called the scuba divers model. In this model, the oceanographer straps on a scuba tank and other associated gear, dives in the water and explores the ocean directly. This is a qualitatively different experience and has historically yielded a different type of result. A third model we might call the aquarium model. The scuba divers model obviously has the handicap of being encumbered with clumsy equipment and such distracting concerns as running out of air. The data analytic analog has the same kind of concerns with wearing a tethering helmet apparatus with limited visual resolution. The aquarium model is an intermediate stage and is analogous to visiting an aquarium with a 40 foot glass wall. We have implemented a 20 foot stereoscopic projection system which will allow the data analyst to view the data through a very large window unencumbered by requirements to wear physical hardware.

Conceptually what we wished to develop is easy to grasp. We proposed a human-computer interface system which will allow the human operator control of his or her motion through the virtual data world simply by movement of an instrumented hand or controller. The binocular visual and binaural audio headgear will also be equipped to sense position and motion. The computations to update views are, in principle, straightforward three-dimensional Euclidean geometry computations. The basic limitations tend to be the accuracy and resolution of the input devices and the throughput of the rendered images. Systems that use a high-powered graphics workstation for each eye still run below 30 frames a second for complex images.

The popular version of immersive technology typically is built around a helmet system which incorporates stereoscopic video, binaural audio and position and motion sensors. The technological weakness of these systems is two-fold. Because of the close proximity of eyes and other sensitive organs, the use of CRT displays seems inadvisable. Most helmet systems have been based on liquid crystal technology which has severe limitations on visual resolution. A second issue for such systems is the latency of response to the motion sensing. There are numerous anecdotal reports on nausea induced by the latency. Both of these problems are technological problems which can be expected to be improved in the future. We have somewhat preferred the stereoscopic projection system as an immersive technology because it is currently capable of a much higher resolution, and, using a six degree of freedom controller, capable of steering the image with much less latency. The effect is that the room essentially becomes a vehicle with a portal to the virtual environment. Both of these immersive technologies offer positive and negative aspects, and both will be explored in the context of this present proposal. Indeed, we envision the situation that both can be used simultaneously by different individuals in the same virtual environment.

Results of the Project

This project had numerous positive results, some anticipated when the proposal was written and some not anticipated. Results generally fall into three categories: immersive methods results, visualization results and general statistical results. The papers and technical reports listed below indicate the extent of our productivity. Individual titles convey a sense of what was accomplished in each of these three categories. Some results

particularly worth mentioned include the use of the our visualization technique called Mode Forest applied to ARL/SLAD ballistic vulnerability data and the visualization techniques we called Saturation Brushing applied to the classification of NSWCC/DD chemical warfare data.

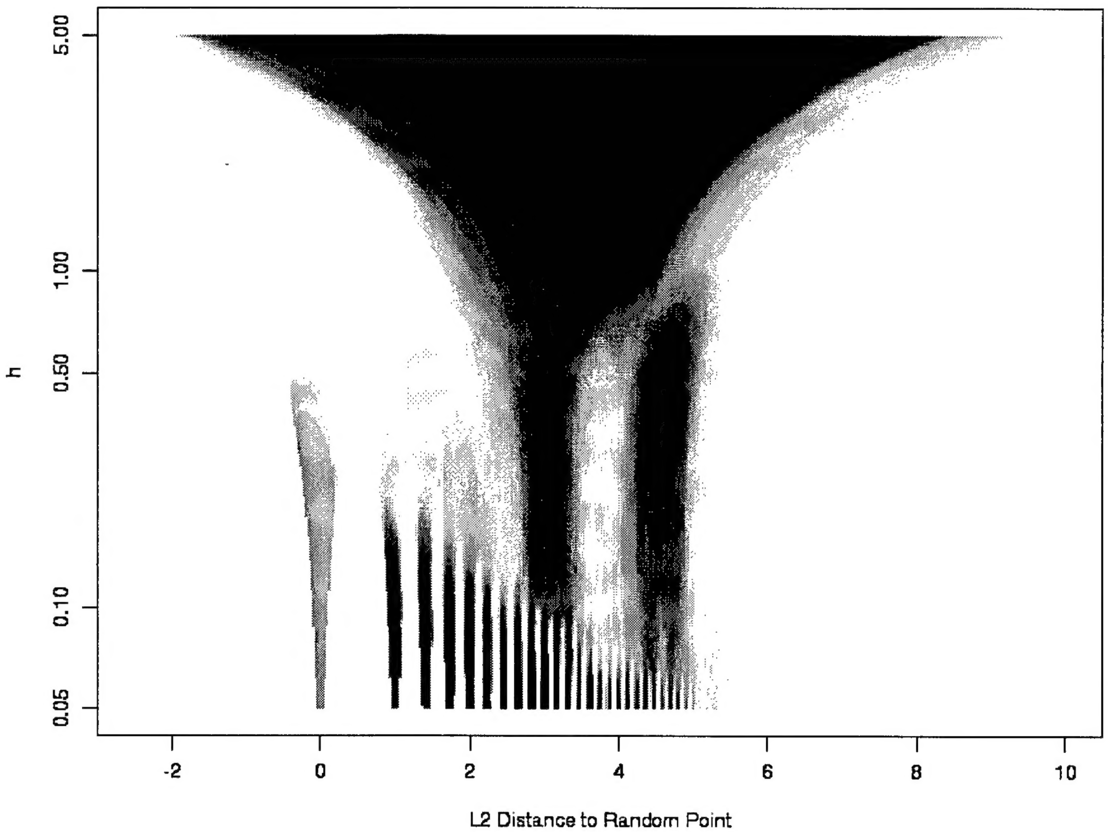

Figure 1. Mode forest visualizing the ARL/SLAD ballistic vulnerability data for the M1 Abrams tank. The two dark regions to the right of the image around 3 and 5 indicate two major clusters while the lighter region around 2 on the abscissa is a smaller cluster. If printing resolution is adequate, there are two subclusters in each major cluster.

The ballistic vulnerability data consists of the investigation of effect of behind armor debris on 67 subsystems of an M1 Abrams tank simulation. Congressional requirements for live fire tests create a dilemma for the U.S. Army. Tanks are obviously extremely expensive targets so that few are used in live fire settings. Because the sample space for binary data on 67 subsystems has 2^{67} items in it, it is unreasonable to expect that the live fire tests will coincide with the outcomes of simulation. The dilemma is thus that it is argued the simulations are no good because they do not correspond to the live fire tests. Our suggestion was to cluster the 67 dimensional data to see if the live fire tests fall into one of the clusters. This strategy has proven very successful because in fact the damage mechanisms do cluster into two main clusters and one minor cluster with two subclusters in each of the major clusters. The graphic illustrating this is included as Figure 1.

The NSWC/DD chemical warfare data involves the measurement of intensity levels of color patches on treated paper in each of 13 spectral bands. The treated paper is

intended to respond to both chemical and biological agents and the nature of the spectral response is indicative of the type of agent. While the 13 dimensional vector is a good discriminator, computations involving the full vector are too slow for a real-time response. It is desired to find simple discriminators that can be computed real-time for an early warning response. The graphic for this is given in Figure 2.

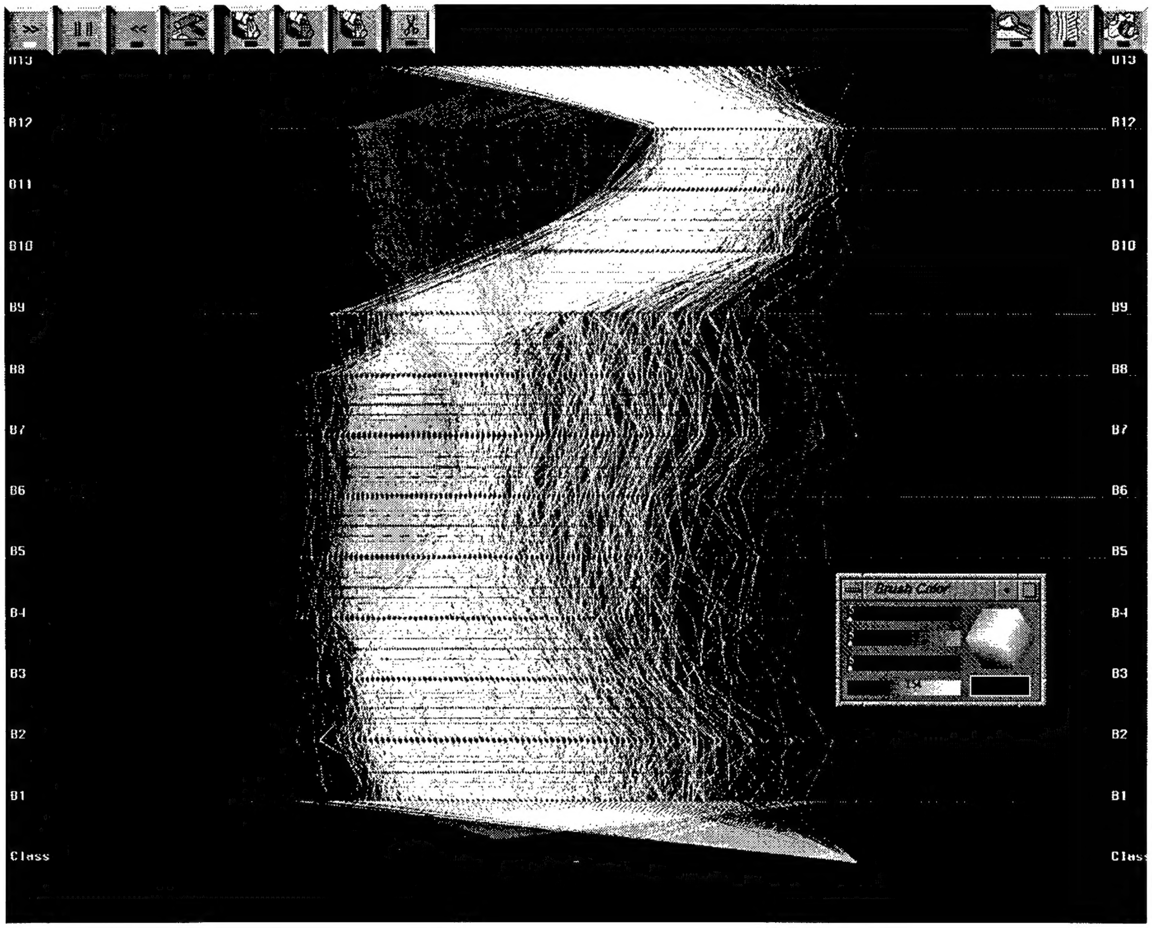

Figure 2. Desaturated parallel coordinate plot of the chemical warfare data indicating that spectral bands B9 and B10 alone are sufficient to discrimate among the three classes of chem-bio agents and even provide discrimation of subclusters with in one type.

This type of graphic allows us to conclude that spectral band B9 and B10 together form discriminators which are as effective as all thirteen bands taken as a whole. With reduced dimensionality the discrimination procedure could be executed real time.

Technical Reports Produced under DAAH04-94-G-0267

TR 143. Edward J. Wegman, Qiang Luo and Jim X. Chen, Immersive Methods for Exploratory Analysis, September, 1997

TR 141. Jim X. Chen and Edward J. Wegman, Real-Time Simulation of Dust Behaviors Generated by Fast Traveling Vehicles, September, 1997

TR 140. Wendy L. Poston, Edward J. Wegman and O. Thomas Holland, Ultrasonic Imaging of Cast Ductile Iron Projectiles, September, 1997

TR 139. Michael C. Minnotte, David J. Marchette and Edward J. Wegman, New Terrain in the Mode Forest, September, 1997

TR 138. David J. Marchette and Edward J. Wegman, Finding Modes with the Filtered Kernel Estimator, September, 1997

TR 136. Don R. Faxon, ***Computational Algorithm for Generalized Nonparametric Function Estimation***, (Ph.D. Dissertation), January, 1997

TR 134. Shumei Wei, ***Image Noise Reduction using Wavelets***, (Ph.D. Dissertation), January, 1997

TR 133. Shan-chuan Li and Edward J. Wegman, Nonparametric Density Estimation using Wavelets and the Method of Sieves, December, 1996. Published, ***Computing Science and Statistics***, 28(1997)641-646.

TR 132. Leonard B. Hearne and Edward J. Wegman, Multivariate Nonparametric Density Estimation using Random Linear Tessellations, December, 1996

TR 131. Winston C. Chow, Edward J. Wegman and Muhammad K. Habib, On the Existence of Some Integrals with respect to Fractional Gaussian Noise, December, 1996

TR 130. Wendy L. Poston, Edward J. Wegman and Jeffrey L. Solka, Robust Estimation of Multivariate Location and Scale based on the Effective Independence Distribution, December, 1996

TR 129. Michael C. Minnotte, David J. Marchette and Edward J. Wegman, The Bumpy Road to the Mode Forest, December, 1996 To appear, ***Journal of Computational and Graphical Statistics***.

TR 128. Edward J. Wegman, Daniel B. Carr, R. Duane King, John J. Miller, Wendy L. Poston, Jeffrey L. Solka and John Wallin, Statistical Software, Siftware and Astronomy, May, 1996, Published with discussion, ***in Statistical Challenges in Modern Astronomy II***, (Babu, G. J. and Feigelson, E. D., eds.) New York: Springer-Verlag, (1997)185-206.

TR 127. Celesta Ball and Edward J. Wegman, Geometric Modeling of Vehicle Paths and Confidence Regions, May, 1996. Published, ***Computing Science and Statistics***, 28(1997)287-292.

TR 126. Ehsan S. Soofi, Information Theoretic Regression Methods, April, 1996.

TR 124. Edward J. Wegman and Qiang Luo, High Dimensional Clustering Using Parallel Coordinates and the Grand Tour, April, 1996. Published ***Computing Science and Statistics***, 28, 352-360, 1997.

TR 123. Edward J. Wegman, Jeffrey L. Solka and Wendy L. Poston, Immersive Methods for Mine Warfare, April, 1996. Published ***MASEVR '95: Proceedings of the Second International Conference on the Military Applications of Synthetic Environments and Virtual Reality***, 203-218, 1996.

TR 120. Barnabas Takacs, Harry Wechsler, and Edward J. Wegman, A Model of Active Perception and its Implementation on the Intel Paragon XP/S, August, 1995.

TR 119. Wendy L. Poston, Edward J. Wegman, and Jeffrey L. Solka, A Parallel Algorithm for Subset Selection, August, 1995. Published ***Journal of Statistical Computation and Simulation***, 60(1998)1-17.

TR 114. Jeffrey L. Solka, Wendy L. Poston and Edward J. Wegman, A New Visualization Technique to Study the Time Evolution of Finite and Adaptive Mixture Estimators, December, 1994, Published ***Journal of Computational and Graphical Statistics***, 4(3)(1995)180-198.

TR 110. Edward J. Wegman, Huge Data Sets and the Frontiers of Computational Feasibility, November, 1994. Published ***Journal of Computational and Graphical Statistics***, 4(4)(1995)281-295.

TR 109. Leonard B. Hearne and Edward J. Wegman, Fast Multidimensional Density Estimation based on Random-width Bins, October, 1994. Published in ***Computing Science and Statistics***, 26(1994)150-155.

TR 106. Wendy L. Poston, Edward J. Wegman, Carey E. Priebe and Jeffrey L. Solka, A Contribution to the Theory of Robust Estimation of Multivariate Location and Shape: EID, October, 1994. Published as "A deterministic method for robust estimation of multivariate location and shape," ***Journal of Computational and Graphical Statistics***, 6(3)(1997)300-313

TR 105. Jeffrey L. Solka, Edward J. Wegman, Carey E. Priebe, Wendy L. Poston and George W. Rogers, A Method to Determine the Structure of an Unknown Mixture Using the Akaike Information Criterion and the Bootstrap, October, 1994.

TR 102. Julia Corbin Fauntleroy and Edward J. Wegman, Parallelizing Locally-Weighted Regression, October, 1994. Published ***Computing Science and Statistics***, 27(1996)327-331.

TR 100. Edward J. Wegman and Qiang Luo, Visualizing Densities, October, 1994. PDF Version.

Published Papers Produced under DAAH04-94-G-0267

Takacs, B., Wechsler, H. and Wegman, E. (1994) "Parallel simulation of an active vision model," ***Proceedings of the Intel Supercomputer Users Group Conference***, 202-207.

Hearne, L. B. and Wegman, E. J. (1994) "Fast multidimensional density estimation based on random-width bins," ***Computing Science and Statistics***, 26, 150-155

Sullivan, M. and Wegman, E. J. (1995) "Correlation estimators based on simple nonlinear transformations," ***IEEE Transactions on Signal Processing***, 43(6), 1438-1444

Sullivan, M. and Wegman, E. J. (1995) "Estimating spectral correlations with simple nonlinear transformations," ***IEEE Transactions on Signal Processing***, 43(6), 1525-1526

Solka, J. L., Wegman, E. J. and Poston, W. L. (1995) "A new visualization technique to study the time evolution of finite and adaptive mixture estimators," ***Journal of Computational and Graphical Statistics***, 4(3), 180-198.

Wegman, E. J. (1995) "Huge data sets and the frontiers of computational feasibility," ***Journal of Computational and Graphical Statistics***, 4(4), 281-295

Wegman, E. J., Poston, W. L. and Solka, J. L. (1996) "Immersive methods for mine warfare," ***MASEVR '95: Proceedings of the Second International Conference on the Military Applications of Synthetic Environments and Virtual Reality***, 203-218

Wegman, E. J., Poston, W. L. and Solka, J. L. (1996) "Wavelets and nonparametric function estimation," in ***Research Developments in Statistics and Probability***. (Brunner, E. and Denker, M., eds.), Utrecht: VSP, 257-274

Solka, J. L., Poston, W. L., Marchette, D. J. and Wegman, E. J. (1996) "Massive data sets in Navy problems," in ***Massive Data Sets: Proceedings of a Workshop***, Washington, D. C.: National Academy Press, 157-167

Wegman, E. J., Le, H. T., Poston, W. L. and Solka, J. L. (1997) "Moments and wavelets in signal estimation," in ***Statistics of Quality***, (Ghosh, S., Schucany, W. R. and Smith, W. B., eds.) New York: Marcel-Dekker, 253-274. Also a short unrefereed version in ***Moments and Signal Processing,*** (Purdue, P. and Solomon, H., eds.), Monterey, CA: Naval Postgraduate School, 270-294, 1992

Sullivan, M. and Wegman, E. (1997) "A normalized correlation estimator for complex data based on a quadruplex transformation," ***IEEE Signal Processing Letters***, 4(1), 26-28

Takacs, B., Wechsler, H. and Wegman, E. J. (1997) "Geometric modeling methods for facial landmark detection and recognition," ***Computing Science and Statistics***, 28, 278-286

Ball, C. G. and Wegman, E. J. (1997) "Geometric modeling of vehicle paths and confidence regions," ***Computing Science and Statistics***, 28, 287-292

Wegman, E. J. and Luo, Q. (1997) "High dimensional clustering using parallel coordinates and the grand tour," ***Computing Science and Statistics***, 28, 352-360

Wallet, B. C., Rogers, G. and Wegman, E. J. (1997) "A mixed measure formulation of the EM algorithm for huge data set applications," ***Computing Science and Statistics***, 28, 492-497

Wallet, B. C., Marchette, D., Solka, J. and Wegman, E. J. (1997) "A genetic algorithm for best subset selection in linear regression," ***Computing Science and Statistics***, 28, 545-550

Solka, J., Wegman, E., Poston, W. and Wallet, B. (1997) "A new iterative adaptive mixtures type estimator," ***Computing Science and Statistics***, 28, 573-578

Li, S.-C. and Wegman, E. J. (1997) "Nonparametric density estimation using wavelets and the method of sieves," ***Computing Science and Statistics***, 28, 641-646

Marchette, D. and Wegman, E. J. (1997) "The filtered mode tree," ***Journal of Computational and Graphical Statistics***, 6(2), 143-159

Wegman, E. J., Daniel B. Carr, R. Duane King, John J. Miller, Wendy L. Poston, Jeffrey L. Solka, and John Wallin (1997) "Statistical software, siftware and astronomy (with discussion)," in ***Statistical Challenges in Modern Astronomy II*** (Babu, G. J. and Feigelson, E. D., eds.) New York: Springer-Verlag, 185-206

Solka, J. L., Wegman, E. J., Rogers, G. W., and Poston, W. L. (1997) "Parallel coordinate plot analysis of polarimetric NASA/JPL AIRSAR imagery," ***Automatic Target Recognition VII - Proceedings of SPIE,*** 3069, 175-184

Poston W. L., Wegman, E. J., Priebe, C. E. and Solka, J. L. (1997) "A deterministic method for robust estimation of multivariate location and shape," ***Journal of Computational and Graphical Statistics***, 6(3), 300-313

Wegman, E. J. (1998) "Parallel coordinate and parallel coordinate density plots," ***Encyclopedia of Statistical Sciences***, Update Volume 2, (Kotz, S., Read, C. B. and Banks, D. L., eds.), 518-525+color plates

Poston, W. L., Wegman, E. J., and Solka, J. L. (1998) "A parallel algorithm for subset selection," ***Journal of Statistical Computation and Simulation***, 60, 1-17

Wegman, E. J., Luo, Q. Chen, J. X. (1998) "Immersive methods for exploratory analysis," ***Computing Science and Statistics***, 29(1), 206-214

Marchette, D. J. and Wegman, E. J. (1998) "Finding modes with the filtered kernel," ***Computing Science and Statistics***, 29(1), 498-507, 1998.

Poston, W. L., Wegman, E. J. and Holland, O. T. (1998) "Ultrasonic imaging of cast ductile iron projectiles," ***Computing Science and Statistics***, 29(1), 292-298

Minnotte, M. C., Marchette, D. J. and Wegman, E. J. (1998) "New terrain in the mode forest," ***Computing Science and Statistics***, 29(1), 473-477

Wallet, B. J., Wegman, E. J., and Marchette, D. J. (1998) "Evolutionary subspace pursuit," ***Computing Science and Statistics***, 29(1), 402-406

Minnotte, M. C., Marchette, D. J. and Wegman, E. J. (1998) "The bumpy road to the mode forest," ***Journal of Computational and Graphical Statistics***, 7(2), 239-251